Social Desirability Tendency and Endorsement of Items in a Forced-Choice Inventory

ABRAHAM I. FELIPE
University of the Philippines

Research on what is now called test-taking behaviors has shown that responses on verbal measuring instruments are often influenced by certain properties of items. One of these properties is the social desirability of endorsing the items, to which Edwards (1953) called attention in his now classic paper. Some self-descriptive statements are socially desirable to ascribe to oneself, while others are not. The degree to which behavior is judged is to be socially desirable "in other people" is highly predictive of the probability of self-ascription.

However, individuals would probably vary in the tendency to endorse self-descriptive statements of a given degree of social desirability. This variation among individuals would probably partly reflect personality differences in the tendency to present oneself to others in socially desirable ways, that is, the SD tendency (Edwards, 1957). Indeed, individuals do seem to differ in the weight they assign to considerations of social desirability in performing many kinds of behavior.

The ability to predict some behaviors, therefore, might be increased by information about the strength of the SD tendency. Previous studies suggest that ascribing to oneself certain inventory statements is one type of behavior predictable from SD measures. The primary purpose of the present study is to extend the implication of these studies, namely, to find out whether information about individual SD tendencies can help predict even the choice between statements arranged in a two-option forced-choice format. In forced choice tests, the subjects (Ss) task is to select their answers from a given list— the familiar multiple-choice question being an example. In the present study, Ss were made to answer on a two-option forced-choice test, i.e., on every item they had to select their answer from only two given alternatives, to each of which was associated a quantitative estimate of its social desirability. Responding in this type of a choice situation was chosen to show the relevance of the SD scale being developed to some measurement techniques in social science research.

The secondary purpose of the study is to validate an experimental SD scale. A scale to measure this response tendency can help social psychological researchers increase their capacity to control some the extraneous factors that influence responding in studies involving human subjects. Many respond in experimental situations in such a way as to only give a socially desirable picture of a self, without regard to their "true" response. In the absence of a way to identify such individuals, experimenters can only randomize the

Originally Published in Philippine Journal of Psychology (1969), Vol. 2, No. 2, pp. 12–18.

assignment of Ss in their experimental conditions. This often leads to running larger numbers of Ss than would otherwise be necessary if they could only control for some of their Ss' relevant personality characteristics, such as the SD tendency. Thus, the SD scale can help refine the procedures in some social psychological studies involving human Ss. It is to be noted the validation of an SD scale implies, in effect, the development of a measure of a personality scale among the Filipinos, for the SD tendency can be conceived of as a dimension of personality (Edwards, 1957; Fordyce, 1956; Marlowe & Crowne, 1964; Merrill & Heathers, 1956).

METHOD

The Experimental UP-SD Scale

From a pool of 101 statements descriptive of behavior that were derived from American-made scales of SD (Edwards, 1967; Marlowe & Crowne, 1964) and the F, K. L. and Manifest Anxiety Scales of the MMPI which were apparently related to SD (Edwards, 1957; Fordyce, 1956; Meehl & Hathaway, 1946; Wheeler, Little & Lehner, 1951), an experimental pool of 32 statements were selected following three steps. First, statements were selected whose social desirability scale-values (SD-SVs) spread out over the whole social desirability continuum. Each SD-SV was a numerical value, derived by successive interval scaling (Edwards, 1952), representing the degree to which the behavior referred to in a statement was perceived by a sample (113 male and female UP students) to be socially desirable "in other people." The first step in item selection, which reduced the item pool from 101 to 49, therefore avoided the exclusions of items of behavior that were only moderately socially desirable. Second, items were next eliminated when less than nine out of ten judges (all UP students) agreed that it would be socially desirable (or undesirable) "in other people" to have such behaviors. This step in item selection further reduced the set to 35 items. After adding 10 other potentially good items to the 35 equalize the number of keyed True and keyed False items, the final 32 item experimental scale was derived by identifying through item-analyses those that contribute most to the total score.

Scoring of the UP-SD Scale

S answered each item either True (the statement applied to him) or False (the statement did not apply to him). When judges agreed that the behavior referred to in a statement was not socially desirable in other people, that statement was keyed True (a score for SD was given if S answered True on that statement). Out of the 32 experimental items, 16 were keyed True and 16 False. S's SD score was the number of items he answered in the keyed direction.

Subjects

The UP-SD Scale was given to 819 students in psychology classes in the University of the Philippines. The SD scores of these students were distributed and the distribution divided into three more or less equal parts. The upper third (Highs) were assumed to

have high SD tendency, the lowest third (Lows) were assumed to have low SD tendency, and the middle third (Moderates) were assumed to have moderate SD tendency.

A total of 94 subjects was run in the present study, 40 Highs (16 males and 24 females), 27 Moderates (7 males and 20 females), and 27 Lows (7 males and 20 females).

Ss were given a 76-item forced-choice inventory. Each of these 76 items consisted of a pair of statements of the form usually found in personality inventories, and the Ss' task was to endorse one of the pair, i.e., to choose which of the two statements in each pair was more descriptive of them. If they could not decide between the two statements, they could check a third alternative, Undecided. The behaviors referred to in the 152 statements in the 76-item questionnaire were previously scaled also (using 58 male and 65 female students in elementary psychology in the University of the Philippines) by the method of successive intervals in terms of how socially desirable they were "in other people", thus generating an Sd-SV for each statement. For example, Statement A and Statement B of Item #28 had SD-SVs of 3.35 and 3.36 respectively, the corresponding SD-SVs of Statement A and Statement B of Item #61 were 3.60 and 1.71. By taking the difference in SD-SVs of each paired statement (scale separations) one derived an estimate of how much more socially desirable is one behavior in comparison with its pair. In the above examples, the scale separations for items #28 and #61 are .01 and 1.89, respectively. It was assumed that the greater the scale separation, the more socially desirable was one of the paired statements relative to the other.

Of the 76-items in the forced-choice inventory, 19 items had scale separations from .01 to .29, another 19 items had from .41 to .68 separations, another 19 had from .82 to 1.14 separations and the last 19 had from 1.20 to 1.90 separations. In other words, depending on the magnitude of scale separations of the statements, the 76-items can be divided into four blocks of 19 items each. Arrangement of these items in the questionnaire was random with respect to scale separation magnitude.

Dependent Variables

Two dependent variables were used: (a) the number of Undecided answers, and (b) the number of times the more desirable (MD) statement of a pair was endorsed.

Predictions in this study pertain to how Ss would choose among the response alternatives in the inventory, assuming the validity of the UP-SD scale.

Predictions Involving the Undecided Responses Alternatives

(a) When scale separations are small, Highs would have more Undecided answers than Lows' but when scale separations are large, Highs would have less Undecided answers than Lows. Highs are assumed to be more concerned than Lows. Therefore, when Highs would have to choose between two equally desirable alternatives, they would find it more difficult than Lows to decide; hence, they would tend to answer Undecided. On the other hand, when one of the alternatives is clearly more socially desirable, they would be more ready than Lows to make a choice. Hence, they would tend to give fewer Undecided answers. Among Moderates, the number of Undecided

answers was expected to be between the Highs' and the Lows' at all levels of scale separation. In analyses of variance terms, the first prediction is an interaction between SD scores x Magnitude of Scale Separation.

(b) More Undecided answers would be given by both Highs and Lows when scale separations are small than when they are large. In analyses of variance terms, this is a predicted main effect due to scale separations.

Predictions Involving the Endorsement of the More Desirable (MD)

Statement:

(a) When scale separations are large, Highs would endorse MD statements more frequently than would Lows since cues for responding in socially desirable ways would then be less ambiguous. However, when scale separations are small, these cues would be ambiguous and, therefore, Highs would have more difficulty in making a socially desirable choice. Hence they might not differ from Lows in the number of MD statements they would endorse. Moderates' answers were again expected to lie between Highs' and Lows.' In analyses of variance terms, an interaction between SD scores x magnitude of scale separations was predicted.

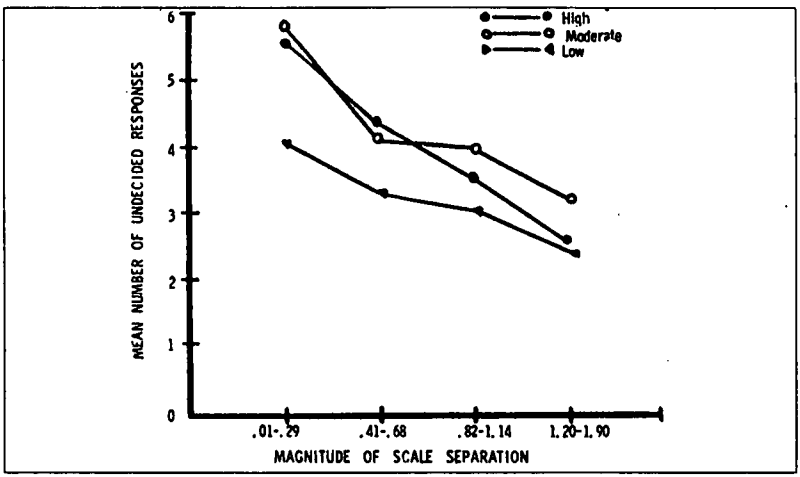
(b) Endorsement of MD statements would increase with scale separation. In analyses of variance terms, this is a predicted main effect due to scale separations.

RESULTS

Number of Undecided Responses

Figure 1 gives the mean number of Undecided responses given by Ss for the four blocks of items. The summary of the analysis of variance of Undecided responses is

Figure 1. Mean number of undecided responses as a function of scale separation



given in Table 1. Magnitude of scale separations is significantly but inversely related to the number of Undecided given, as predicted. A more detailed analysis indicates that 97 percent of the variance explained by scale separations are accounted for by a linear function—that is, as scale separation increases in magnitude, the number of Undecided decreases monotonically.

Table 1. Analysis of variance of undecided responses

Source	*df*	*MS*	*F*
Between Ss:			
A (SD Scores)	2	17.58	
B (sex)	1	0	
AB	2	45.78	
Error	88	27.89	
Within Ss:			
C Scale Separation) 16.41***	3	39.89	
a. linear 47.74***	1	116.02	
b. quadratic	1	0.77	
c. cubic	1	2.83	
AC	6	7.59	3.12***
a. linear	2	20.79	8.56***
b. quadratic	2	1.64	
c. cubic	2	0.76	
BC	3	2.28	
ABC	6	6.08	2.50*
Error	264	2.43	

*p < .05
**p < .01
***p < .001

The predicted interaction between SD scores and magnitude of scale separation was also significant. Straight lines seem to explain most of the variations in the answers of Highs, Moderates and Lows, but the significant linear scale separation x SD scores interaction indicates that the best fitting linear functions for these three groups have different slopes. In other words, the number of Undecided given by Highs, Moderates and Lows tended to differ as a function of scale separation. This was, of course, predicted. The pattern of means as well as the reported t-tests, however, indicated only partial confirmation of the predictions. Highs and Lows responded partly as predicted; the former gave significantly more Undecided responses on the two item-blocks with the smallest scale separations $t = 2.76$, $df = 45$, $p < .01$ on the .01-.29 item-block and $t = 2.51$, $df = 45$, $p. < .01$ on the 41.-68 item-block). They did not differ significantly in other two item-blocks with larger scale separations, on both of which Highs gave numerically

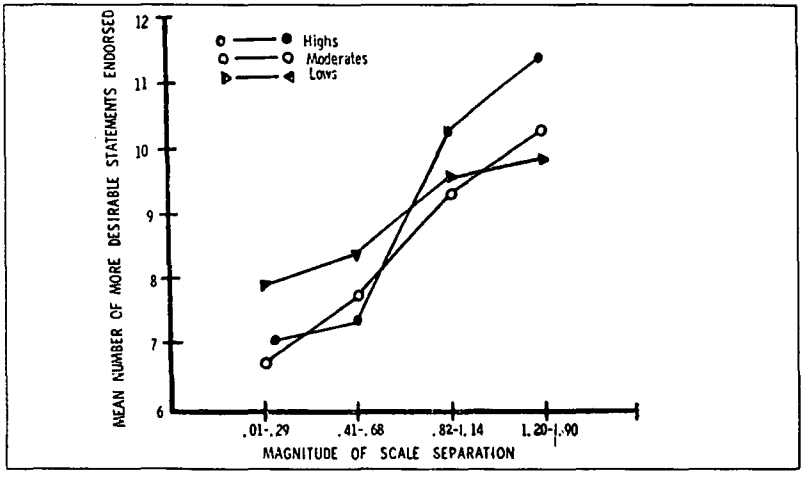
but not significantly more Undecided, contrary to prediction. The responses of the Moderates were not as predicted also. Compared with Lows, they gave more Undecided s across all item-blocks, but significantly more only on the block with smallest scale separations ($t = 2.40$, $df = 52$, $p < .01$). Compared with Highs, they also gave more Undecided s on 3 of the 4 items-blocks, although no difference between them was significant.

Table 1 also shows an unexpected but significant triple interaction which seems to be largely associated with sex differences among Moderates when they respond on items with .01–.28 scale separations but not on the other three blocks of items: male Moderates gave less Undecided on this item-block than either male Highs or male Lows, whereas female Moderates gave more Undecided than either female Highs or female Lows. No explanation for this complex finding will be forwarded in this paper.

Number of MD Statements Endorsed

Figure 2 shows the mean number of MD statements endorsed in the different conditions. Except for the responses of the Moderates on two blocks of items (.01–.28 and .82–1.14 scale separations), the pattern was as predicted. The summary of the analysis of variance is given in Table 2. As scale separation between statements increased, there was significantly more endorsement of MD statements . This relationship seems to be mostly linear but the significant cubic trend indicates that an S-shape curve explains a significant portion of the variance. Depending on their SD scores Ss, however, differed in the tendency to endorse the MD statement as a function of increasing scale separation (as indicated by the significant interaction between SD scores and Scale separation). The linear Scale Separation x SD Scores interaction indicates that the linear best-fitting lines for the Highs, Moderates and Lows have different slopes.

Figure 2. More desirable statements endorsed as a function of scale separation



On t-tests, Highs and Moderates did not differ from each other on the two blocks of items with the smallest separations whereas both, compared with Lows, endorsed significantly fewer MD statements (on the .01-.28 item-block, $t = 2.04$, $df = 65$, $p < .01$ and $t = 2.17$, $df = 52$, $p < .01$ for Highs vs. Lows and Moderates vs. Lows respectively; the corresponding values on the .41–.68 item block are $t = 2.38$, $df = 65$, $p < .01$ and $t = 2.19$, $df = 52$, $p < .01$). On the other hand, for the two blocks of items with the largest separations, Lows and Moderates did not differ from each other at all. However, compared to Highs, both endorsed fewer MD statements and significantly so on the 1.20–1.90 item-block ($t = 2.50$, $df = 65$, $p < .01$, fr Highs vs. Moderates; and $t = 3.41$, $df = 65$, $p < .01$ for Highs vs. Lows).

Table 2. Analysis of variance of more socially desirable statements that were endorsed

Source	*df*	*MS*	*F*
Between Ss:			
A (SD Scores)	2	6.86	
B (sex)	1	26.10	
AB	2	29.67	
Error	88	12.9	
Within Ss:			
S (Scale Separation)	3	159.03	36.59***
a. linear	1	427.81	98.43***
b. quadratic	1	0.06	
c. cubic	1	58.66	13.50**
AC	6	11.23	2.58*
a. linear	2	29.67	6.83*
b. quadratic	2	1.38	
c. cubic	2	2.22	
BC	3	4.22	
ABC	6	1.74	
Error	264	4.35	

*$p < .05$
**$p < .01$
***$p < .001$

DISCUSSION

It appears that the UP-SD scale, interpreted as an instrument to measure the SD tendency, can help predict how Ss would answer a two-option forced-choice test involving alternatives which differ in their social desirability loadings. Apparently, in this kind of task, individuals with high scores on the UP-SD Scale differ in a predictable way from individuals with low scores. In comparison with the latter, high scorers found it more difficult to decide between alternatives which were closely equated in their SD-SVs and they endorsed more MD type of statements when the alternatives were widely separated in their SD-SVs.

There are, of course, some unpredicted observations from the present study that suggest difficulties for the UP-SD Scale. Two of these findings can be disposed of as possible procedural artifacts, a third is more serious. The possible artifacts are: (a) the significantly fewer MD type of statements endorsed by Highs on items with small scale-separations, and (b) the failure to get significantly fewer Undecided among Highs on items with large scale separations. The first was very likely an artifact of the response measures since for items with small scale-separations Highs gave more Undecided than Lows, and in the inventory the number of Undecided given is inversely related to the total number of MD statements that could be endorsed . The two response measures were not independent of each other in the present study since everytime a statement was endorsed, there was one statement less wherein Undecided could be given as a response, and vice-versa. The second could have been an artifact of the magnitude of scale separations employed. In the 1.20–1.90 item-block, here Highs were predicted to give significantly fewer Undecided than Lows, the prediction might have not confirmed because the scale separations between alternatives might not be large enough since they were mostly on the lower side of the range. Using larger scale separations between alternatives would be advisable in a replication of this study.

The more serious unpredicted finding has to do with the Moderates who, contrary to predictions, did not behave as a differentiable group but, on the other hand, generally performed more like the Highs. This implies that the linearly interpreted scores on the UP-SD scale does not reflect a corresponding (linear) psychological dimension, otherwise Moderates would have at least scored between Highs and Lows. While this implication is common and never disastrous in psychological measurement, it is not desirable.

The difficulty to differentiate Moderates from Highs, while both can be differentiated from Lows, implies that they are the Lows whose behaviors can be said to predictable from UP-SD scores. The hardly differentiable Moderates and Highs comprise two-thirds of the distribution and can therefore define how "most people" behave. The Lows, however, behave differently from "most people," and it is in this sense that one may conclude that the UP-SD scale helps predict only the behavior being predicted is endorsement of forced-choice items. Experiments to study the predictive validity of the scale using other criterion behaviors are presently underway.

The UP-SD Scale and the findings in the present study have relevance to some social science research in the Philippines. Some verbal instruments in the social sciences may be answered in such a way as to produce only certain socially desirable effects on others. Individuals can be expected to differ in their readiness to take advantage of the opportunity afforded by these instruments to produce such social effects. It seems that the UP-SD Scale can be used to some extent as a measure of that readiness. When groups are to be compared on an instrument that is vulnerable to SD response sets, it might be useful to form them first by matching Ss in terms of their SD scores.

REFERENCES

Edwards, A.L. (1952). The scaling of stimuli by the method of successive intervals. *Journal of Applied Psychology,* 36, 118–122.

Edwards, A .L. (1957). *The social desirability variable in personality assessment and research.* Drydne Press, New York.

Fordyce, W. E. (1956). Social desirability in the MMPI. *Journal of Consulting Psychology, 20,* 171–175.

Mrlowe, D. P. & Crowne, D. (1964). *The Approval Motive.* John Wiley and Sons, New York.

Meehl, R. M. & Hathaway, S. R. (1946). The K factor as suppressor variable in the MMPI. *Journal of Applied Psychology, 30,* 525–564.

Merrill, R. M. & Heathers, Louise B. (1956). The relation of the MMPI to the Edwards Personal Preference Schedule on a college counseling center sample. *Journal of Consulting Psychology, 20,* 310–314.

Wheeler, W. M., Little, K. B. & Lehner, G. F. J. (1951). The internal structure of the MMPI, *Journal of Consulting Psychology, 15,* 134–141.